

```
from sklearn.datasets import load_diabetes
data = load_diabetes()
```

```
print(data.DESCR)
```

```
➞ .. _diabetes_dataset:
```

Diabetes dataset

Ten baseline variables, age, sex, body mass index, average blood pressure, and six blood serum measurements were obtained for each of n = 442 diabetes patients, as well as the response of interest, a quantitative measure of disease progression one year after baseline.

****Data Set Characteristics:****

:Number of Instances: 442

:Number of Attributes: First 10 columns are numeric predictive values

:Target: Column 11 is a quantitative measure of disease progression one year after baseline

:Attribute Information:

- age age in years
- sex
- bmi body mass index
- bp average blood pressure
- s1 tc, total serum cholesterol
- s2 ldl, low-density lipoproteins
- s3 hdl, high-density lipoproteins
- s4 tch, total cholesterol / HDL
- s5 ltg, possibly log of serum triglycerides level
- s6 glu, blood sugar level

Note: Each of these 10 feature variables have been mean centered and scaled by the standard

Source URL:

<https://www4.stat.ncsu.edu/~boos/var.select/diabetes.html>

For more information see:

Bradley Efron, Trevor Hastie, Iain Johnstone and Robert Tibshirani (2004) "Least Angle Regression" (https://web.stanford.edu/~hastie/Papers/LARS/LeastAngle_2002.pdf)

◀

```
x = data.data
y = data.target
```

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2)
```

```
from sklearn.linear_model import Ridge
r = Ridge(alpha = 0.0001)
```

```
r.fit(x_train,y_train)
```



▼ Ridge
Ridge(alpha=0.0001)

```
y_pred = r.predict(x_test)
```

```
from sklearn.metrics import mean_squared_error, r2_score  
import numpy as np
```

```
rmse = np.sqrt(mean_squared_error(y_test, y_pred))  
r2 = r2_score(y_test, y_pred)
```

```
print("RMSE:", rmse)  
print("R2 Score:", r2)
```



RMSE: 53.953806860863224
R2 Score: 0.4693576165297858